

Transfer Learning Analysis on Tuberculosis Classification Using MobileNetV2 Architecture Based on Chest X-Ray Images

Ali Samsul Latupono^{1*}, Majid Rahardi^{2**}

* Teknik Informatika, Universitas Amikom Yogyakarta
alisamsul665@students.amikom.ac.id¹, majid@amikom.ac.id²

Article Info

Article history:

Received 2025-10-15

Revised 2025-11-07

Accepted 2025-11-15

Keyword:

*Tuberculosis,
MobileNetV2,
Deep Learning,
Transfer Learning,
Chest X-ray,
Image Classification.*

ABSTRACT

Tuberculosis(TBC) remains a major global health issue, with millions of new cases reported annually. Early and accurate diagnosis is essential, but manual interpretation of chest X-ray(CXR) images is limited by subjectivity and resource constrains. This study applies the MobileNetV2 architecture using transfer learning to classify tuberculosis from CXR images. The publicly available Tuberculosis Chest X-ray dataset containing 4200 images was divided into training (70%), validation (15%), and testing (15%). The pretrained MobileNetV2 model on ImageNet was used as the base network, with additional classification layers and training through the Adam optimizer and early stopping. The model achieved a validation accuracy above 99.84% after the second epoch maintained stable performance. Once the test set, model reached 99.84% accuracy, with precision 99.53% and recall 99.90% for the tuberculosis class. The result demonstrate that the transfer learning with MobileNetV2 provides a fast, efficient, and highly accurate method for tuberculosis detection. This model show potential for integration into Computer-Aided Diagnosis (CAD) system in low resource clinical settings.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Tuberkolosis merupakan penyakit menular yang disebabkan oleh *Mycobacterium tuberculosis* dan menjadi penyebab utama kematian akibat oenyakit infeksi. Menurut laporan *World Health Organization (WHO)* tahun 2022, terdapat 10.6 juta kasus baru dan 1.6 juta kematian akibat TBC di seluruh dunia [1]. Radiografi dada (CXR) merupakan alat utama dalam proses diagnosis karena ketersediaannya yang luas dan biaya yang relative rendah. Namun, interpretasi manual oleh radiolog masih menghadapi keterbatasan waktu, kelelahan visual, dan perbedaan Tingkat pengalaman [2].

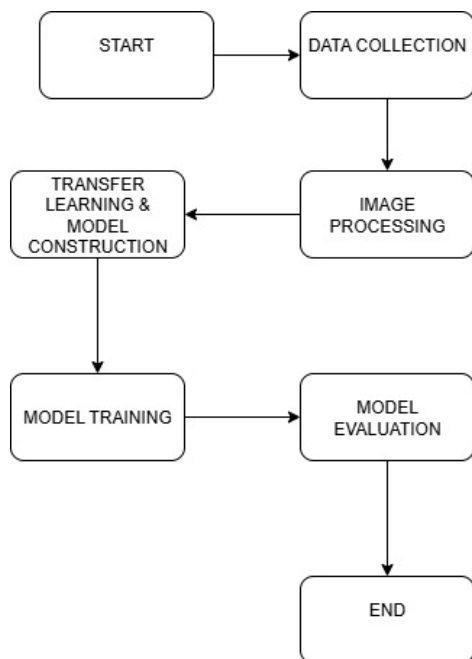
Perkembangan *artificial intelligence (AI)* dan *deep learning*, terutama dengan *Convulutional Neural Network (CNN)*, memberikan hasil yang menjanjikan dalam analisis medis [3]. Metode *transfer learning* memungkinkan pemanfaatan model yang telah dilatih pada dataset besar seperti ImageNet untuk digunakan kembali pada domain media, sehingga dapat menghasilkan performa tinggi dengan data terbatas [4].

Penelitian terdahulu menunjukkan bahwa arsitektur ringan seperti MobileNetV2 mampu mencapai akurasi tinggi dengan kebutuhan komputasi rendah, menjadikannya ideal untuk lingkungan klinis [5]. Model lain seperti DenseNet dan Vision Transformer (ViT) juga banyak digunakan untuk analisis citra dada [6], [7]. Studi lapangan di Filipina menunjukkan bahwa system AI-CAD berbasis CXR memiliki sensitivitas tinggi dalam mendeteksi TBC di dunia nyata [8].

Berdasarkan hasil tersebut, penelitian ini bertujuan menerapkan dan mengevaluasi MobileNetV2 dengan *transfer learning* untuk klasifikasi TBC dari *image CXR*. Model ini diharapkan dapat menjadi Solusi cepat, efisien, dan akurat bagi fasilitas Kesehatan dengan sumber daya terbatas.

II. METODE

Penelitian ini dilakukan dalam beberapa tahap, seperti yang ditunjukkan pada Gambar 1. Metode yang digunakan meliputi pengumpulan data, pemrosesan gambar, pembuatan model, pelatihan model, evaluasi model



Gambar 1. Alur Penelitian

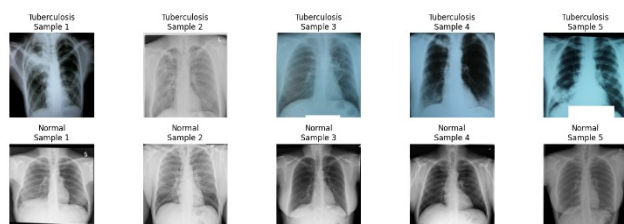
A. Data Collection

Penelitian ini menggunakan *Tuberculosis Chest X-ray* Dataset yang berisi 4200 citra rontgen yang terdiri dari dua kelas yaitu Tuberculosis dan Normal. Data dibagi menjadi 70% pelatihan, 15% validasi, dan 15% pengujian dengan propors seimbang [9].

TABEL I
DISTRIBUSI DATASET

CLASS	TOTAL IMAGE
TUBERCULOSIS	700
NORMAL	3500

Untuk memberikan visual dari data yang digunakan, gambar 2 menampilkan contoh citra dari setiap kelas dalam dataset.



Gambar 2. Sampel Citra Dataset

B. Image Processing

Tahap ini dilakukan untuk memastikan seluruh citra memiliki kualitas dan ukuran yang seragam sebelum digunakan pada proses pelatihan model. Setiap citra diubah ukurannya menjadi 224x224 piksel agar sesuai dengan format input MobileNetV2. Nilai piksel dinormalisasi ke rentang 0-1 menggunakan operasi *rescaling* ($1./255$) untuk

meningkatkan stabilitas dan efisiensi proses pelatihan. Proses augmentasi diterapkan melalui rotasi acak, flipping horizontal, dan zoom ringan untuk menambahkan variasi data dan mengurangi risiko overfitting.

Label kelas (Normal dan TBC) dikonversi ke format *one-hot encoding* agar sesuai dengan fungsi *loss categorical_crossentropy*. Pembagian data dilakukan secara otomatis menggunakan fungsi *image_dataset_from_directory()* TensorFlow dengan pengacakan terkontrol dan rasio validasi 15%. Tahap ini memastikan data memiliki ukuran, skala intensitas, dan distribusi label yang konsisten sehingga model dapat melakukan pembelajaran secara optimal terhadap fitur yang relevan.

C. Model Architecture

Model yang digunakan Adalah MobileNetV2 dengan bobot awal dari ImageNet. Lapisan dasar dibekukan (*frozen*) agar berfungsi sebagai pengekstrak fitur, kemudian ditambahkan beberapa lapisan baru, yaitu:

- Lapisan GlobalAveragePooling2D
 - Lapisan Dense(128 neuron, aktivasi ReLU)
 - Lapisan Dropout (rate 0,5) untuk mencegah *overfitting*
 - Lapisan Dense output (2 neuron, aktivasi softmax)
- Model dilatih dengan optimizer Adam dan *Loss Function categorical_crossentropy*. Teknik *Early Stopping* digunakan untuk menghentikan pelatihan Ketika *validation loss* berhenti menurun [10].

Hyperparameter. Proses pelatihan model dilakukan dengan serangkaian hyperparameter yang telah ditetapkan untuk memastikan keterulangan eksperimen. Kami menggunakan optimizer Adam, sebuah algoritma optimasi adaptif, dengan *learning rate* bawaan (default) sebesar 0.001. *learning rate* ini mengontrol laju pembaruan *weights* model. Fungsi kerugian (loss function) yang digunakan adalah 'categorical_crossentropy' yang umum digunakan untuk tugas klasifikasi.

Model dilatih menggunakan batch size sebesar 32, yang berarti *model* memperbarui *weights*-nya setiap 32 gambar. Untuk mencegah *overfitting*, kami menerapkan lapisan Dropout setelah lapisan GlobalAveragePooling2D dengan *rate* 0.5 (50%). Pelatihan juga menerapkan *callback* Early Stopping yang memonitor *validation loss*, yang menghentikan proses pelatihan secara otomatis pada *epoch* ke-8 ketika model tidak lagi menunjukkan peningkatan performa.

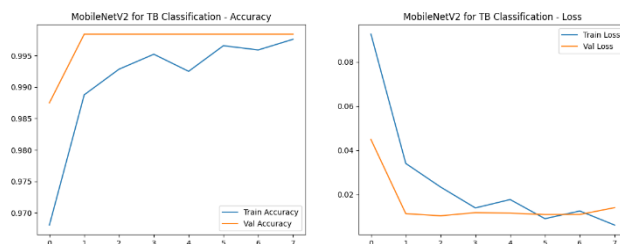
D. Model Evaluation

Evaluasi model dilakukan dengan menggunakan metrik akurasi, presisi, recall, F1-Score, dan *confusion matrix*. Selain itu, hasil dibandingkan dengan model lain seperti DenseNet dan ViT untuk menilai peningkatan kinerja [11], [12].

III. HASIL DAN PEMBAHASAN

A. Training Performance

Selama pelatihan, modelss menunjukkan konvergensi cepat dalam delapan epoch. Grafik pada gambar 3 memperlihatkan bahwa akurasi pelatihan dan validasi meningkat signifikan sejak epoch pertama, kemudian mencapai stabilitas di atas 99% pada epoch kedua. Nilai *Validation loss* turun tajam hingga dibawah 0.01 dan tetap rendah hingga akhir pelatihan, menandakan bahwa model tidak mengalami *overfitting* meskipun jumlah epoch relatif kecil.



Gambar 3. Training Model

B. Evaluation Data

Evaluasi pada data uji menggunakan 620 citra menunjukkan performa yang sangat baik. Model mencapai akurasi sebesar 99,84% dengan nilai presisi rata-sarat 99.53%, recall 99.90%, dan F1-Score 99.72%. Hasil ini menandakan bahwa model memiliki kemampuan klasifikasi yang sangat akurat serta seimbang dalam mendeteksi citra positif Tuberkulosis maupun citra normal. Nilai metrik yang tinggi menunjukkan bahwa model mampu mengenali pola citra dengan baik tanpa adanya perbedaan signifikan antara data latih dan data uji.

Evaluasi Model MobileNetv2 untuk Klasifikasi Tuberkulosis	
Akurasi	: 99.84%
Presisi	: 99.53%
Recall	: 99.9%
F1-Score	: 99.72%

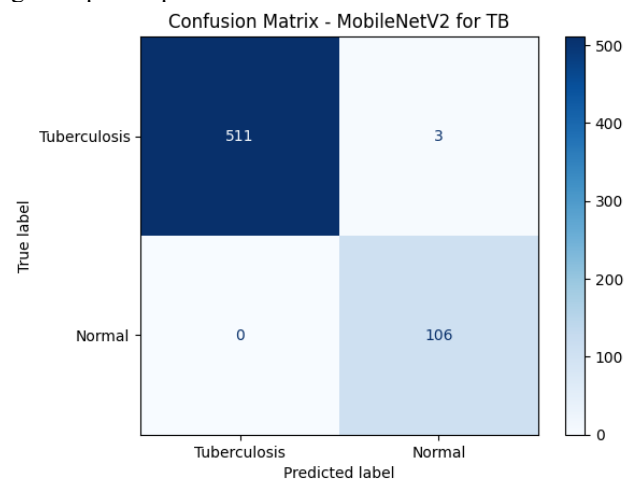
Gambar 4. Hasil

Meskipun hasil akurasi 99.84% yang dicapai model tampak sangat tinggi, kami menyadari bahwa akurasi setinggi ini pada dataset medis publik memerlukan analisis kritis. Seperti yang telah dicatat sebagai keterbatasan, ada kemungkinan bahwa faktor-faktor seperti *data leakage* (kebocoran data) atau *overfitting* harus dapat berkontribusi. Walaupun grafik *validation loss* (gambar 3) tampak stabil, tingginya performa mungkin juga dipengaruhi oleh karakteristik dataset yang digunakan. Dataset ini memiliki ketidakseimbangan kelas yang signifikan (3500 Normal vs 700 TBC), dan ada kemungkinan bahwa citra dalam dataset latih dan uji memiliki kesamaan visual yang tinggi (misalnya berasal dari sumber akuisisi yang sama), yang membuat tugas klasifikasi menjadi lebih mudah bagi model daripada jika

dihadapkan pada data klinis di dunia nyata yang lebih bervariasi.

C. Confusion Matrix Analysis

Confusion matrix pada gambar 5 menunjukkan distribusi prediksi model. Dari total 620 citra, sebanyak 511 citra TBC dan 106 citra normal berhasil diklarifikasikan dengan benar. Hanya terdapat tiga kesalahan klasifikasi, yang semuanya merupakan *false negative* pada kelas TBC. Tidak ditemukan *false positive* pada kelas Normal, menunjukkan Tingkat keandalan tinggi model dalam menghindari kesalahan diagnosis positif palsu.



Gambar 5. Confusion Matrix

Analisis *Confusion Matrix* (Gambar 5) menunjukkan performa model yang sangat baik dalam menghindari *False Positive* (0 kasus). Hal ini sangat bermanfaat secara klinis karena mengurangi risiko diagnosis positif palsu dan pemeriksaan lanjutan yang tidak perlu. Namun, kami mengidentifikasi adanya tiga kasus *False Negative*. Meskipun angka ini kecil implikasi klinisnya sangat signifikan. *False Negative* berarti model salah mengklasifikasikan pasien TBC sebagai 'Normal', yang dalam praktis klinisnya nyata dapat menyebabkan kegagalan diagnosis dini dan penundaan pengobatan. Hal ini menyoroti area di mana sensitivitas model masih perlu ditingkatkan sebelum implementasi klinis.

D. Comparison with other Studies

Hasil ini menunjukkan peningkatan performa relatif terhadap beberapa studi sebelumnya yang memakai arsitektur lain, misalnya studi dengan DenseNet yang melaporkan akurasi ~94% [4], serta model ringan lain seperti LightTBNNet yang melaporkan akurasi di kisaran rendah 90-an pada dataset tertentu [12]. Model *hybrid* CNN-ViT pada beberapa studi juga menunjukkan akurasi tinggi namun memerlukan sumber daya komputasi lebih besar [13].

Jika dibandingkan dengan arsitektur lain, performa MobileNetV2 dalam studi ini melampaui beberapa studi sebelumnya yang menggunakan DenseNet (~94%). Namun, keunggulan utama MobileNetV2 tidak hanya terletak pada akurasinya, tetapi pada efisiensi komputasionalnya.

Arsitektur *lightweight* (ringan) ini dirancang khusus untuk perangkat *mobile* dan *edge computing*. Hal ini menjadikannya kandidat ideal untuk implementasi sistem *Computer-Aided Diagnosis* (CAD) di fasilitas kesehatan dengan sumber daya terbatas atau untuk aplikasi diagnostik berbasis ponsel, di mana model yang lebih berat seperti VIT mungkin tidak praktis karena keterbatasan daya komputasi perangkat keras.

Namun, beberapa temuan perlu diperhatikan untuk memberikan gambaran yang lebih objektif. Pertama, keberadaan sejumlah kecil *false negative* menunjukkan bahwa model belum sepenuhnya sensitive terhadap semua variasi citra TBC, sehingga masih ada potensi kasus yang tidak terdeteksi. Kedua, proses validasi dilakukan menggunakan satu dataset publik sehingga belum dapat memastikan kemampuan generalisasi model terhadap populasi pasien dan perangkat radiografi yang berbeda. Ketiga, Tingkat akurasi yang sangat tinggi dapat mengindikasikan adanya faktor yang memengaruhi keaslian hasil, seperti ketidakseimbangan distribusi data atau kebocoran informasi antarset. Oleh karena itu, diperlukan evaluasi lanjutan dengan dataset independen dan pemeriksaan ulang pada proses pembagian serta pra-pemrosesan data untuk memastikan reliabilitas model.

Analisis *Confusion Matrix* (Gambar 5) menunjukkan performa model yang sangat baik dalam menghindari *False Positive* (0 kasus). Hal ini sangat bermanfaat secara klinis, namun, kami mengidentifikasi adanya tiga kasus *False Negative* yang implikasi klinisnya sangat signifikan sehingga menyebabkan kegagalan diagnosis dini. Untuk mitigasi, disarankan pengujian lintas dataset, augmentasi lebih bervariasi, serta eksperimen ensemble dan interpretabilitas (mis. Grad-CAM) untuk memeriksa area yang memengaruhi Keputusan model [14].

IV. KESIMPULAN

Penelitian ini menunjukkan bahwa penerapan metode *transfer learning* dengan arsitektur MobileNetV2 efektif untuk melakukan klasifikasi Tuberkulosis (TBC) berdasarkan citra rontgen dada. Melalui tahap *image preprocessing*, pelatihan model, dan evaluasi berbasis metrik akurasi, presisi, dan *recall*, diperoleh performa yang sangat tinggi dengan akurasi validasi mencapai 99.5% dan akurasi uji sekitar 99.8%. Hasil ini mengindikasikan bahwa MobileNetV2 mampu mengekstrak fitur citra paru dengan efisien dan menghasilkan prediksi yang sangat akurat, bahkan pada dataset berukuran sedang.

Selain menunjukkan performa tinggi, hasil penelitian ini juga menegaskan potensi model ringan seperti MobileNetV2 untuk implementasikan pada sistem diagnosis berbantu komputer (*Computer Aided Diagnosis*) di fasilitas Kesehatan dengan keterbatasan sumber daya. Berdasarkan hasil evaluasi confusion matrix dan laporan klasifikasi, model ini memiliki kemampuan yang stabil dalam membedakan citra Normal dan TBC dengan Tingkat kesalahan yang sangat rendah. Hal ini memperlihatkan bahwa pendekatan *transfer learning* mampu

mengoptimalkan model pra-latih untuk aplikasi medis tanpa memerlukan pelatihan dari awal.

Meskipun demikian, beberapa aspek masih memerlukan pengujian lebih lanjut untuk memastikan rehabilitas model. Validasi yang dilakukan masih terbatas pada satu dataset publik sehingga generalisasi terhadap variasi populasi dan perangkat radiografi belum dapat dipastikan sepenuhnya. Oleh karena itu, penelitian lanjutan disarankan untuk menggunakan dataset multi-sumber, menerapkan Teknik *fine-tuning* lanjutan, serta menambahkan analisis interpretabilitas seperti Grad-CAM agar sistem berbasis AI dapat dipercaya dan digunakan secara aman dalam konteks klinis.

DAFTAR PUSTAKA

- [1] "Global Tuberculosis report 2022," 2022. [Online]. Available: <http://apps.who.int/bookorders>.
- [2] M. Nel, Z. Franckling-Smith, T. Pillay, S. Andronikou, and H. J. Zar, "Chest Imaging for Pulmonary TB—An Update," Feb. 01, 2022, *MDPI*. doi: 10.3390/pathogens11020161.
- [3] Z. Zhang, F. Cui, C. Cao, Q. Wang, and Q. Zou, "Single-cell RNA analysis reveals the potential risk of organ-specific cell types vulnerable to SARS-CoV-2 infections," *Comput Biol Med*, vol. 140, Jan. 2022, doi: 10.1016/j.combiomed.2021.105092.
- [4] O. Illah and A. Olaitan, "Updates on HPV Vaccination," Jan. 01, 2023, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/diagnostics13020243.
- [5] D. Musinguzi *et al.*, "PaliGemma-CXR: a multi-task multimodal model for tuberculosis chest X-ray interpretation," *BMC Artificial Intelligence*, vol. 1, no. 1, p. 7, Aug. 2025, doi: 10.1186/s44398-025-00008-3.
- [6] M. Tanaka, S. Shiota, and H. Kiya, "A detection method of operated fake-images using robust hashing," *J Imaging*, vol. 7, no. 8, Aug. 2021, doi: 10.3390/jimaging7080134.
- [7] Y. Hadhoud *et al.*, "From Binary to Multi-Class Classification: A Two-Step Hybrid CNN-ViT Model for Chest Disease Classification Based on X-Ray Images," *Diagnostics*, vol. 14, no. 23, Dec. 2024, doi: 10.3390/diagnostics14232754.
- [8] N. Marquez *et al.*, "Performance of chest X-ray with computer-aided detection powered by deep learning-based artificial intelligence for tuberculosis presumptive identification during case finding in the Philippines," *BMC Global and Public Health*, vol. 3, no. 1, Dec. 2025, doi: 10.1186/s44263-025-00198-y.
- [9] K. K. Sampath, A. Arumugam, and G. Jull, "The role of 3D digital applications in Manual Therapy Education – A scoping review," Dec. 30, 2021, doi: 10.21203/rs.3.rs-1178829/v1.
- [10] L. Kaewwilai *et al.*, "Development and evaluation of an artificial intelligence (AI) -assisted chest x-ray diagnostic system for detecting, diagnosing, and monitoring tuberculosis," *Glob Transit*, vol. 7, pp. 87–93, Jan. 2025, doi: 10.1016/j.glt.2025.02.005.
- [11] V. Sharma, Nillmani, S. K. Gupta, and K. K. Shukla, "Deep learning models for tuberculosis detection and infected region visualization in chest X-ray images," *Intelligent Medicine*, vol. 4, no. 2, pp. 104–113, May 2024, doi: 10.1016/j.imed.2023.06.001.
- [12] D. Capellán-Martín, J. J. Gómez-Valverde, D. Bermejo-Peláez, and M. J. Ledesma-Carbayo, "A Lightweight, Rapid and Efficient Deep Convolutional Network for Chest X-Ray Tuberculosis Detection," Sep. 2023, doi: 10.1109/ISBI53787.2023.10230500.
- [13] S. A. El-Ghany, M. Elmozy, M. A. Mahmood, and A. A. Abd El-Aziz, "A Robust Tuberculosis Diagnosis Using Chest X-Rays Based on a Hybrid Vision Transformer and Principal Component Analysis," *Diagnostics*, vol. 14, no. 23, Dec. 2024, doi: 10.3390/diagnostics14232736.
- [14] B. U. Maheswari *et al.*, "Explainable deep-neural-network supported scheme for tuberculosis detection from chest radiographs," *BMC Med Imaging*, vol. 24, no. 1, Dec. 2024, doi: 10.1186/s12880-024-01202-x.